

The integrated anatomy practical paper: a robust assessment method for anatomy education today

Article (Accepted Version)

Smith, Claire F and McManus, Bruce (2015) The integrated anatomy practical paper: a robust assessment method for anatomy education today. *Anatomical Sciences Education*, 8 (1). pp. 63-73. ISSN 1935-9780

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/48343/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

ASE-13-0130.R2

Descriptive article

The Integrated Anatomy Practical Paper (IAPP): A robust assessment method for anatomy education today

Claire F. Smith^{1,2}, Bruce McManus²

¹Department of Anatomy, Brighton and Sussex Medical School, University of
Sussex, Brighton, East Sussex, United Kingdom

²Medical Education, Faculty of Medicine, University of Southampton, Southampton,
United Kingdom

Running Title: The Integrated Anatomy Practical Paper

Correspondence to: Dr. Claire France Smith, Medical School Building, Biology Road,
University of Sussex, Falmer, BN1 9PX, United Kingdom. E-mail:
medicineanatomyDA@bsms.ac.uk

ABSTRACT

Assessing anatomy in a way that tests higher cognitive domains and clinical application is not always straightforward. The old 'spotter' examination has been criticized for only testing low level 'identify' knowledge, while other assessment modalities such as multiple choice questions (MCQs) do not reflect the three dimensional and application nature of clinical anatomy. Medical curricula are frequently integrated and subject specific examinations do not reflect the case based, spiral, integrative nature of the curricula. The Integrated Anatomy Practical Paper (IAPP) is a hybrid of the old 'spotter' and an Objective Structured Clinical Examination (OSCE) but it demonstrates how higher levels of taxonomy can be assessed, together with clinical features and integrates well with other disciplines. Importantly, the IAPP has shown to be reliable and practical to administer. Data gathered from the Bachelor of Medicine five year program over two academic years for four IAPP examinations, each being 40 minutes with (K = 60 items) based on 440 students revealed consistently strong reliability coefficients (Cronbach alpha) of up to 0.923. Applying Blooms taxonomy to questions has shown a marked shift resulting in an increase in the complexity level being tested; between 2009 and 2013 a reduction of 26% in the number of low level 'remember knowledge' domain questions was noted with up to an increase of 15% in 'understanding' domain and 12% increase in the 'applying' knowledge domain. Our findings highlight that it is possible to test, based in a laboratory, anatomy knowledge and application that is integrated and fit for practice.

Key words: gross anatomy education, medical education, laboratory examination, practical examination, spotter examination, assessment methods, OSPE, integrated curricula, learning approach, knowledge level.

INTRODUCTION

Anatomy assessment has traditionally involved modalities such as a spotter examination, written examinations and oral examinations (viva voice). With changing curriculum tides (Drake, 1999) anatomy assessment is often integrated and in many institutions does not involve a practical component. This is reflected in the demise of the traditional spotter examination despite recommendations that a practical examination is more suitable for testing anatomical knowledge than a written examination (Rowland et al., 2011). The traditional spotter examination where students have to identify a pinned structure has been a part of anatomy assessment for some considerable time despite being criticized for testing low levels of knowledge (Yaqinuddin et al., 2013). Basic factual understanding is a key component to learning and as described by Smith and colleagues (Smith et al., 2014) when applied to the notion of threshold concepts (Meyer and Land 2003) the basic building blocks, which can be classed as 'identify,' are important in the anatomy learning journey.

A variant of the spotter examination is the 'steeplechase method' (Chirculescu et al., 2007) where students have to identify a structure and then state simply its action or function. Anatomy can also be assessed using an Objectively Structured Practical Examination (OSPE) which frequently includes clinical skills. An OSPE traditionally involves a procedure based station and a question based station (Yaqinuddin et al., 2013). In the model provided by Yaqinuddin and colleagues the OSPE presents each question with a clinical vignette (Yaqinuddin et al., 2013). Anatomy has also traditionally been assessed by an oral examination. Concerns over bias and reliability have reduced the role of the oral examinations in UK medical schools. In a

study (Duffield and Spencer 2002) which asked medical students at Newcastle University only 46% of students felt the oral examinations were fair. However, in the United States Clough and Lehr recommended the use of oral examinations to emphasize clinical applications, spatial relationships, nomenclature and functions. (Clough and Lehr 1996). More recently, oral examinations have been recommended as a more suitable alternative to the spotter examination for physical therapy students (Fabrizio, 2013). A range of justifications have been proposed for using a spotter assessment, they include the following:

- Spotter examinations use real human specimens which is important for clinical practice,
- Spotter examinations test three dimensional spatial understanding,
- Spotter examinations test students understanding of relations of structures to each other which is important for intervention diagnostics and surgery,
- Spotter examinations can test ability to differentiate similar structures e.g. nerves and blood vessels which is important in clinical practice,
- Spotter examinations can be used to test for an appreciation of anatomical variation which is important in clinical imaging diagnostics and intervention.

It could be asked though, are the justifications above really met by a question asking for an individual to identify or 'spot' a structure?

Both the spotter and steeplechase examinations reflect a beach 'flag and sandcastle' approach, where a pin is placed in a structure with a flag on top of it with a number or letter and the student has to identify what the pin is in i.e., the sandcastle;

however it can be difficult in the spotter or steeplechase setting to do anything more, especially anything integrated or requiring a high level of complexity.

Over recent years there has been enhancement of alignment between the curriculum and the material tested. This is frequently referred to as 'blueprinting' (Crossley et al., 2002) so each examination question is mapped to the learning outcome. It is widely understood that assessment drives learning (Biggs and Tang, 2011) and that assessment has to be constructively aligned to the curriculum (Biggs and Tang, 2011). In anatomy education it is understood that the approach to learning influences the assessment outcome in anatomy (Smith and Mathias, 2010). A surface approach was significantly related to a low achievement level and a deep or strategic approach a high attainment level. Hence the need to ensure assessment rewards a deep approach to learning and is reflective of the purpose of learning i.e. clinical practice. In a further study that involved alumni it was found that approach to learning also influenced the career choice for doctors (Smith and Mathias, 2011). The importance of the purpose of learning in context has been highlighted by Böckers al. (2014). Where participation in a clinical elective designed to highlight applied clinical anatomy resulted in a positive assessment trend. A study by Bergman and colleagues highlighted several important findings; that curricula style does not correlate with students' knowledge and that good test performance is related to total teaching time (Bergman et al., 2008).

Context

The Centre for Learning Anatomical Sciences (CLAS) at the University of Southampton, previously the Department of Human Morphology was established in

1972. CLAS has always had a laboratory environment with a range of resources including human cadavers. In 2009 CLAS moved to a hospital site and into a new facility. As reflected by Heylings curricula had shifted and along with 57% of the UK, Southampton has a systems based curricula (Heylings, 2002).

ANATOMY TEACHING

Anatomy is taught as part of a five year Bachelor of Medicine Program. The Bachelor of Medicine Program is designed as a spiral case led curriculum. A spiral curriculum is designed so that students revisit topics as they develop (Bruner, 1960). For example in the first year there is a Nervous and Locomotor 1 (NLM1) module that spirals into the year two Nervous and Locomotor 2 module (NLM2), which spirals into the relevant clinical specialties. Students study anatomy as part of system based modules in years one and two, with typically six modules containing anatomy. Figure 1 provides an overview of years one and two of the Bachelor of Medicine program. Year's three to five of the program involve clinical placements and a Bachelor of Science research component. The curriculum for anatomy is in line with the Core Curriculum for Medical Students as is recommended by the General Medical Council Tomorrows Doctors (McHanwell et al., 2007).

Within the case led modules (Figure 1) a typical structure for anatomy is a series of lectures that occur before laboratory based prosection sessions. The prosection sessions are task based and introduced with a short tutor led demonstration aided by AV capture. Students are guided by workbooks to utilize a range of prosections, plastinated specimens, models, pathology potted specimens, osteology and e-learning material, via touch screens or tablet. Students are free to work how they

wish, although in practice they naturally work with their peers in small groups. Figure 2 illustrates students working in the laboratory with their workbooks. Each module has an anatomy workbook that covers the anatomy relevant to the module which is a mixture of regional and system-based. The workbook contains text book style text integrated with tasks that students complete during the session. There are clinical sections describing examples of anatomy as applied to practice. The workbook also contains twenty questions per practical to help students ascertain if they are working to the required level. For example: *Having considered the psoas muscle and its fascia what might be the clinical symptoms of a right sided psoas abscess?* A number of these workbooks have now been made into electronic fully interactive resources which the students can access in the laboratory through faculty purchased iPADS or their own mobile phones or tablets which have been found to be very popular with students (Cecot et al., 2013).

Two anatomists are available during the class to offer guidance. A typical prosection session involves around thirty embalmed prosections. In addition anatomy teaching also involves timetabled histology sessions and living anatomy sessions. Students are free to visit and use the laboratory for private study during normal working hours. The prosected specimens and cases are selected to support spiral learning. For example in NLM1, the first year module for the nervous and locomotor systems most of the limb anatomy is covered with basic neurological detail, whereas in the second year module covering the same systems (NLM2) the focus is on head and neck anatomy including neuroanatomy. In the Cardiovascular, Respiratory and Renal One module (RCR1) the anatomy related to these systems are covered. In RCR 2 the content is based around the clinical application often using props, for example

central lines on specimens, to reinforce the applied nature of anatomy and the living anatomy session moves to a more applied use of ultrasound. On average a module has thirty anatomy hours devoted to it per student, a total of one hundred and eighty hours over the first two years.

Anatomy Formative Assessment

Throughout a module students are exposed to typically three formative Integrated Anatomy Practical Papers (IAPP) prior to examination (18 in total over year one and two), these are termed IFAs and stand for Identify, Function and Application Stations. The IFAs include questions from histology, pathology, physiology, pharmacology and clinical questions so that students can get used to the integrated format. A typical IFA will cover two prosection teaching sessions and will involve thirty questions set in a similar manner to the IAPP. The IFA offers students immediate feedback on their learning as they can openly browse the questions and answers in their own self-directed study time. The IFA is usually available to students for two weeks in the laboratory. Students can complete the IFA on their own or in pairs. In year two students have been very keen to help faculty set up these IFAs and the setting of them has been reported to be a valuable learning experience in its own right.

The IFA rating system was created as a student friendly version of a taxonomy system such as the revised Blooms taxonomy (Anderson, 2000) so students could be better informed of the level of the questions and to gain further feedback. The questions and answers are coded with an “I”, “F” or “A” rating which clearly shows students in a simple manner an idea of the purpose and difficulty of the question. It is

not always the case but frequently an identify (“I”) question involves a single level of thought. For example; *Identify the structure pinned A. (Answer: pituitary gland)*. The function (“F”) level of questions often involves two stages. For example; *State two hormones which are stored but not made in the structure pinned A. (Answer: oxytocin and ADH)*. In this example a student has had to identify that the structure pinned is the pituitary gland and then answer the physiology based component. For application (“A”), there may be two or three levels involved. For example; *a patient with a non-secreting tumor at A would display what two clinical features? Answers: (any two of: headache, vision loss, particularly loss of peripheral vision, nausea and vomiting)*. In this example the student has to first of all work out what structure A is, then they have to know the surrounding anatomy to think about what structures might be affected and then at the third stage understand what the function are of those structures. This is in contrast to the old spotter where questions were predominantly Identify. This is also in contrast to the Steeplechase method where A and B are linked, for example A might be a pinned structure and B will ask its function.

Anatomy Summative Assessment

Anatomy Spotters were a part of the Bachelor of Medicine year and term examination structure with three spotters per year. The spotter was utilized successfully for many years. The development of a hybrid assessment started to occur with a drive to increase the cognitive level of the assessment. Hence moving away from the ‘spot’ identify style of question. Students reported being surprised when more clinical and functional questions were in the spotter, even though they had been informed of this development and had formative examples. The student

perception was that it would be a flag and sandcastle style examination, limited to anatomy. At the same time in 2007/2008 a curriculum review was establishing a new Year one and two of the Bachelor of Medicine five year program. The key elements of the new curriculum were an increased amount of spiraling of subjects and increased integration of subjects from the first day to the last. The curriculum and anatomy faculty were all keen to keep a practical style of assessment. As such there was no drive to reduce the resource in terms of cost. The anatomy team was more concerned that the examination was aligned and fit for purpose to test competence in understanding and applying anatomy and associated disciplines.

New assessment structure

The curriculum changes resulted in all assessments being further integrated and reflecting the spiral in the curriculum. As a result anatomy was further integrated into other assessment modalities such as MCQ and short note papers. It was therefore not appropriate to have a separate anatomy assessment. However it was recognized and supported by the faculty that there was a need to have an assessment which reflected how students had learnt and applied their anatomy. As a result the new assessment needed to contain the following: integration with other subjects, relate to the application of the knowledge, use a wide range of resources that students have used in their learning journey and be blueprinted to the Learning Outcomes. There was no divide in the curriculum to material being presented as 'normal' in year one and 'pathological' in year two. All modules contain elements of normal and pathological right from the start. Understanding students perceptions we were keen to label the examination correctly and the term Integrated Anatomy Practical Paper (IAPP) was decided.

The summative assessment structure for years one and two is semester based, with assessment periods in January and June. Three examinations are taken, a single best answer examination (MCQ), a written examination and the IAPP. The written examination involves short and long structured questions. Additional requirements for progression include competence in Medicine in Practice and completion of Student Selected Units (Components). Each examination is standard set. Students who are 10% below the pass mark (referred to as the minimum pass mark) have failed that examination and the semester regardless of their performance in the other examinations. Students who fail to reach the pass mark but are above the minimum pass mark may compensate their overall mark if they have higher scores in the others examinations. The examinations are weighted with 40% of the overall mark contributed by each of the written examination and the single best answer and 20% by the IAPP. Students who fail the semester have the opportunity to re-take all components not just the failed examination during the supplementary examination period in August.

INTEGRATED ANATOMY PRACTICAL PAPER DESIGN

The new IAPP comprises of thirty stations each station with an A and B question, each question being worth two points. The IAPP therefore contains sixty units and is worth one hundred and twenty points. In evaluating the amount of time given to each question thirty seconds was decided and hence one minute was allocated per station. After one minute a buzzer sounds and students move to the next station in a clockwise direction. Throughout the IAPP there are a series of resting stations, refer to Figure 3. The decision of how much time was allocated per question was based

on recommendations for Objectively Structured Clinical Examinations and is similar in time to the ninety seconds frequently given to a steeplechase examination (Shaibah and van der Vleuten, 2013). A recent study has suggested that there is no significance difference between a timed and untimed steeplechase (Zhang et al., 2013) suggesting that timing might be more of a practical consideration than a cognitive one. Due to the number of medical students (Table 1) the IAPP runs with two circuits; an inner and an outer one, Figure 3. This does cause some logistical problems in ensuring matching prosections. In order to assess all students three rotations are run. In the last rotation students that qualify for extra time (as determined the University Support Services, for example due to a specific learning difficulty like dyslexia) are permitted an extra seven and a half minutes once the other candidates have left.

The selection of what material goes into the IAPP was undertaken by a matrix which takes into account the number of practical hours students have had in a course on anatomy, histology, pharmacology, pathology and physiology. The examination questions are blueprinted to the learning outcomes, sample questions can be seen in Table 2. Colleagues in other disciplines are contacted and asked to provide possible questions. To deter students remembering and sharing questions with junior colleagues each IAPP was composed of at least 50% new questions. Re –used questions are important as they offer comparison between cohorts.

The process of setting the IAPP involves a day of an academics time, arranging specimens and resources. As shown in Figure 3 a series of tables are dedicated to specimens but this does not mean they are restricted to anatomy. For example on a

sagittal section of a female pelvis specimen an 'A' pin may be placed in the cervix and students may be asked a question around drugs used to dilate the cervix. A 'B' pin could be placed in the same specimen in the uterine tube asking about ectopic pregnancy symptoms. Again this is different to the steepplechase method as 'B' is not reliant on 'A'.

The IAPP is standard set by the academic team to produce a cut point score to determine those who perform well enough and those who do not (Norcini, 2003). The standard setting team involves a mixture of anatomists, scientists in other disciplines e.g. physiologist and at least one clinician. The IAPP standard setting uses a criterion referenced Modified Ebel's method (Ebel, 1979). This method is an item based absolute method where questions are rated into groups according to difficulty (e.g. easy, moderate or difficult). Then panel members make a judgment, for each category, regarding the percentage of questions a borderline student would get correct. These percentages are multiplied by the relative proportion of the total questions that are assigned to each category, and the results for each category are summated, to arrive a final cut score. For example if it is judged that 40% are easy, and that 70% of borderline candidates will get that question correct, while 40% are moderate with 60% getting those correct and finally 10% are difficult with 30% getting that correct, then the cut score is $(0.4 \times 0.7) + (0.4 \times 0.6) + (0.1 \times 0.3) = 0.28 + 0.24 + 0.03 = 0.55$ The cut score is thus 55%. This is modified from the original Ebel's method which also adds a dimension regarding importance for each question. However, we have made the assumption that all questions are of equal importance.

As described by (Cohen-Schotanus and van der Vleuten, 2010) a gold standard does not exist for standard setting. Therefore in 2012 we created a study to determine Ebel's suitability for the IAPP. A potential other method is the Angoff method (Angoff, 1971). While both systems are an absolute method making judgments about individual test items, the Angoff method required the judges to estimate the proportion of borderline candidates who were likely to respond to each question correctly. For each question an average of the individual panel members judgments is calculated then using the average for all questions is calculated and this is the cut score. The Ebel and Angoff methods were conducted simultaneously in 2012 but it appeared that the standard setting process and the makeup of the panel was too expert to use the Angoff method for this type of the examination. For sixty items the Angoff method was tedious and the group being composed of mainly anatomists had high expectations of the students. In this evaluation the Ebel's proved more reliable and easier to administer. The Ebel's pass mark for the IAPP is frequently between 45-55%. In order to validate the panels judgments, a third method is used post hoc; the Hosftree method (De Gruijter, 1985) which asks judges what the maximum and minimum cut score and fail rate should be. Then using a graph showing the distribution of students, these parameters allow a cut score to be interpolated. (Norcini, 2003; Jalili et al., 2011). A photographic record is taken for all stations and shown to the External Examiners.

The marking of the IAPP is straightforward as the mark scheme allows for only a few answers. As good practice, initially ten examination scripts are marked (checked/corrected) and the mark scheme reviewed. Throughout the marking process if answers appear that are worthy of marks these are added to the mark

scheme. Rules for marking are dealt with at the beginning; for example what to do if a student adds in a dermatome level. In this scenario one correct dermatome plus or minus one is usually awarded 1 mark, rather than 2. Rules are also established for students who write their answers in the wrong place for example by applying a 5% penalty to the affected questions. Once the scripts are marked they are electronically scanned and total marks calculated. A moderation meeting then occurs where all fails are double checked and reviewed, together with 10% of the other scripts from a range of marks. Scripts just over or below the pass mark (borderline scripts) are also re-examined and any recommendations or changes are noted for the Board of Examiners.

The Board of Examiners is a validation stage where the External Examiners (two or three) of the program review and ratify the examination results for all assessment components within year one and two prior to publication. Data is also then provided back to the teaching team on how students performed on each question. On the day of the IAPP after all the circuits have run students are allowed in for a thirty minute feedback session. Students can then look at the stations with staff to gain timely feedback on their performance. During the session students are prohibited from making any notes or taking pictures so that a component of the IAPP questions can be used in the future. Frequently only twenty to thirty students use this opportunity as it is late in the afternoon. If the number of students significantly increased the session would be limited to twenty minutes and only a certain number of students would be allowed in, before swapping with the waiting group.

ANALYSIS

To determine if the IAPP is a robust assessment three key features were explored: The validity and reliability, the cognitive level and overall achievement. This evaluation study was initiated by the Taught Programme Assessment Committee in the University of Southampton, as part of a review for program development; it was exempt from requiring Institutional Review Board approval. All data were analyzed using a combination of statistical packages: SPSS (SPSS Inc., Chicago IL), Microsoft Excel (Microsoft Corp., Redmond WA), and 'R' (R Foundation for Statistical Computing, Vienna, Austria).

Validity and Reliability

A sample of two academic years' data from years one and two Bachelor of Medicine five year program were analyzed. This included the IAPPs of each semester (2 a year) covering the following modules: Foundations of Medicine, Nervous and Locomotor 1, Respiratory, Cardiovascular and Renal 1, Gastrointestinal, Nervous and Locomotor 2, Respiratory, Cardiovascular and Renal 2 and Endocrinology and Life Cycle, Table 1 provides details of sample size.

Cronbach Alpha is a measure of reliability which seeks to assess the average degree of inter-item correlation or co-variance and thus has a value from 0 – 1. Put another way it seeks to determine the degree to which items (questions) are consistent with each other in measuring the performance of candidates within the construct, it is sometimes referred to as internal consistency. A value greater than 0.8 suggests reliability is good. The Cronbach alpha for each IAPP is shown in Table 3. All IAPPs show a high level of reliability. Demonstrating reliability is an important first step in demonstrating validity. A reliable assessment is a pre requisite for a valid

assessment, but demonstrating reliability does not in itself also demonstrate validity.

The validity of the IAPP is also supported by:

- Firstly the arguments that have already been presented regarding the construct of the IAPP seeking to test higher cognitive levels. That is that the IAPP measures not only that students can identify a structure (“I”) but that they are also able to integrate that knowledge together with knowledge of other disciplines and describe function (“F”) and apply knowledge (“A”).
- Secondly by selecting and blueprinting the content to sample widely across the integrated course content and learning outcomes.

Cognitive Level

To tease out the cognitive level a rating was applied to each question, based on the revised Blooms taxonomy (Anderson, 2000) where six levels are noted. The levels are as follows: Remembering Knowledge (1), Understanding (2), Applying (3), Analyzing (4), Evaluating (5), and Creating (6), (Table 4). The ratings were applied at the time of standard setting over Semester 4 examinations (2009-2013). The decision on the rating was aided by examining the active verbs within the question. In the old spotter examination data from 2009 shows 72% of questions were rated at the lowest level of 1, compared to 2013 where 46% were rated at the level of 1 ‘remembering knowledge’. In 2009 0% of questions were rated at level 4 ‘Analyzing’, compared to 2013 where 7% of questions were rated at level 4, (refer to Table 4). As can be seen the first shift has been in reconfiguring questions from straightforward identify questions ‘level 1, remembering knowledge’ into higher level functional questions that also tested the students ability to know what the structure was. The second shift was in developing the higher levels and bringing integration in.

Overall Achievement

IAPP results from each Semester were examined from 2009-2013 to explore students achievement. To do this anonymized spotter examination and IAPP results were examined in the following ways: mean, median, \pm SD, number of fails, and interquartile range and are displayed using the Semester 4 IAPPs 2009-13 as an example in the format of a box and whisker plot (Figure 4). Further the data suggests that the differences are not simply the result of one cohort being better or worse than another. There is a significant trend towards lower marks between 2009-2011 when the IAPP was introduced. There are then some significant improvements between 2011 and 2012, although this has been partially lost in 2013 (Figure 5). The differences are seen both within cohorts and between cohorts suggesting it is independent of cohort effects. This suggests these changes perhaps relate to either the difficulty of the examination or to teaching and the students preparedness. It could be hypothesized that the introduction of the IAPP led to a rising of the standard required to pass but then teaching and student preparedness also improved to “rise to the challenge”.

The mean marks for the IAPPs were plotted with error bars representing 1.96 times their standard deviation, Figure 6 (therefore 95% of the cohort achieved a mark within the error bars). It is noted that the variance of scores is roughly similar for each IAPP, suggesting that whilst the standard has risen there still exists a similar range of student ability. The important part of this is that the standard of the whole cohort down to the lower 2.5th centile has also risen. Furthermore, Figure 6 shows the mean in relation to its 95% confidence interval (1.96 times the standard error of

the mean). Therefore comparing semesters 1 & 2 the drop in performance between 2009 and 2010 is statistically significant while comparing 2010 or 11 (after the introduction of the IAPP) with the same semesters in 2012 and 13 the rise is also statistically significant.

DISCUSSION

In order to consider the different components of the IAPP the following discussion has been divided into sections that will explore; the knowledge level, student evaluation, practicalities, advantages and disadvantages and a summary.

Knowledge Level

It was once criticized that spotters test only low level knowledge, the IAPP has overcome this to test to a high cognitive level. It is easy to pin a structure and ask a student to identify it. To enhance the question the same structure can be pinned but the question can be made more difficult and applied by having sequential steps. The first step is the identification of x, the second is the working out that x supplies y and the third part is that a clinical sign of this would be z. As the steps progress the cognitive demand is increasing whilst the time to answer the question remains the same. In thirty seconds this is a tall order and tests not just the identification knowledge but is testing the understanding, together with reasoning and quick thinking skills including decision making. It has been shown that limiting time at anatomy examinations does not adversely affect the student performance (Zhang et al., 2013); although students tell us that the time restraints put a positive pressure on them.

As demonstrated in Table 4 it is not possible to use all of Blooms taxonomy levels and this is appropriate for students in Years one and two of a medical program, the highest level of creating involves 'planning and producing', for example this might be correctly citing a chest drain or completing a drug chart based on correct history taking and diagnosis. Such activities are found in later clinical years. It would also not be appropriate for too few questions to be at level 1, as for medical students in their first and second year it is essential that they grasp the fundamental building blocks. Such building blocks and 'knowing that' may be linked to a surface approach to learning but as (Smith et al., 2014) describes this part of learning is essential in enabling students to get key concepts or threshold concepts (Meyer and Land, 2003). At the same time it is the integration and application of knowledge which helps students adopt a deep approach to learning which is more preferential in the long term (Smith and Mathias, 2010). As students are frequently driven by the assessment, an assessment such as the IAPP can be very powerful in promoting a deep approach to learning anatomy.

In the IAPP questions are designed to be independent, compared to the OSPE example (Yaqinuddin et al., 2013) where part B is interrelated to part A. In the example above a student does not need to know that the hormones ADH and oxytocin are stored in the pituitary in order to know that a non-secreting tumor might exert pressure effects on the optic chiasm. Selecting questions carefully ensures that students are not disadvantaged in part B by an inability to answer part A. Therefore each part is a new opportunity to demonstrate their ability increasing the number of independent items and so also the reliability coefficient of the IAPP.

Student Evaluation

Student feedback regarding all assessment and the preparation for assessment was discussed at module focus groups. Initially the move to the IAPP caused a certain degree of concern for students. Over the first couple of years student perception changed and as the hidden curriculum messages (Hafferty, 1998) were passed from student to student. The hidden curriculum is the material not hidden from students but is the local communications between students that are a result of their education. In this case the concern soon turned into positive feelings of the IAPP. After the first two years students accepted the integrated nature of all of the examinations. Further discussions over recent years have highlighted that the students have found the IAPP in their words to be 'difficult' but in a positive way as they felt it reflected more appropriately how they use their knowledge and also how the quick thinking nature prepares them for OSCEs in later years. For example, they felt that examinations such as Single Best Answer tested their understanding or remembering ability of a single facet but that the IAPP tested more in line with how they might use their knowledge in an integrated manner. Due to the IFAs students continually report feeling supported and prepared for the style of assessment. Students use of the IFA differs with some students reporting on using it once as a 'mock' and others revisiting it several times using it more as a longitudinal learning tool. There have been no difficulties noted or reported by students for whom English is not their first language or who have a specific learning difficulty. The IAPP examination has been undertaken with a range of adjustments for students on crutches or in wheelchairs.

Practicalities

On a practical note the IAPP responds well to enabling a high volume of students to go through a practical examination with its duplicate stations because the stations involve a mixture of radiology, osteology, models, prosections, pathology pots, plastinated specimens and histology pictures it is easy to resource it, easier than the old spotter, on average thirty prosections are required for both circuits to accommodate seventy students. The IAPP is set by one academic member of staff with technical support. On the day of the IAPP University rules require one academic member of staff to be there with two other invigilators. Due to the nature of the laboratory environment the invigilators are the laboratory technicians and they help with the distribution and collection of scripts. The resource list for an IAPP can be used several years in a row as a number of questions can be changed; we recommend a 50/50 split between new and reused questions. Whilst it has been suggested that a Selected Response Format (SRF) could replace Free Response Format (FRF) answers in practical examinations (Shaibah and van der Vleuten, 2013); with the IAPP there is no need for SRF as the answers are short and easy to mark. FRF also avoids the potential downside of SRF in that it does not provide the student with a cue; instead the students must generate the correct answer themselves. The IAPP is digitally added.

Advantages and disadvantages

This paper, like others (Chirculescu et al., 2007; Bergman et al., 2008, Yaqinuddin et al., 2013,) provides a description of the assessment practice based on sound pedagogical evidence but goes on to further providing evidence of the reliability and

validity of the IAPP method. The IAPP has expanded the justifications for using it to include the following:

- It is aligned to integrated learning which underpins the constructive alignment of the curricula
- It is clinically relevant which promotes a deep approach to learning
- It can test students ability to apply clinical knowledge in the diagnostic domain
- It can test students' ability to apply clinical knowledge in the treatment domain.

Whilst the IAPP is currently used as part of the Bachelor of Medicine Program a smaller scale IAPP was utilized for a one off Bachelor of Science module 'Building the Human Body' this worked well and was adapted with the clinical component removed and the modules evolution and developmental aspect added. Whilst not tried in other medical school curricula, it is proposed that an IAPP format would work in curricular including a traditional regional pre-clinical approach and Problem Based Learning approach as well as science based modules. In summary the main advantages and disadvantages are highlighted as:

Advantages:

- The ability to improve students understand and performance in anatomy by rewarding a deep approach to learning.
- Engages and is in line with integrated spiral curricula.
- Designed to test a range of cognitive levels
- Is a good discriminator of students who have reached the required level of competence.

- Deal with a high volume of students
- Relatively quick and easy to administer.

Disadvantages:

- There still remains an inability to handle specimens which does not therefore encompass the tactile aspect of learning anatomy as explained by Fabrizio 2013. This Touch Mediated Perception (Smith 2010) has been shown to be a positive component associated with a deep approach to learning.
- For high student numbers requires duplication of resources to run two examinations simultaneously.

CONCLUSIONS

It is possible to assess a range of taxonomy levels of anatomy knowledge and understanding in an integrated evidence based approach that is valid, reliable and economical / practical in its resource requirements. With evidence showing that the context of learning anatomy does not make a difference (Smith et al., 2014) it is also recommended that the IAPP would be suitable for anatomy assessment in professions allied to medicine.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the work of colleagues in the Centre for Learning Anatomical Sciences and the Taught Program Assessment Committee both led by Dr. Jenny Skidmore who have all contributed to the IAPP, its design and delivery. Our gratitude is expressed to Mrs. Sally Rushworth, assessment team leader for the Faculty of Medicine and her team for their assistance with the examination spread sheets. Thank you to Mrs. Lucy Law for Figures 2 and 3.

NOTES ON CONTRIBUTORS

CLAIRE FRANCE SMITH, B.Sc., P.G.C.E., Ph.D., F.H.E.A., F.A.S., is reader in Anatomy and Head of Anatomy at Brighton and Sussex Medical School, University of Sussex, Brighton East Sussex, United Kingdom. Her research is in understanding the learning experience including: approaches to learning, learning psychometrics and spatial ability. She is also Chair of Education Committee for the Anatomical Society of Great Britain.

BRUCE MCMANUS, B.Sc. M.B.B.S., P.G.C.E., M.R.C.G.P., D.R.C.O.G., D.F.S.R.H., D.G.M., is the faculty lead for assessment analysis and development and principal clinical teaching fellow at the University of Southampton, Faculty of Medicine, Southampton, United Kingdom.

LITERATURE CITED

Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, Raths J, Wittrock MC. 2000. A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. 2nd Ed. Old Tappan, NJ: Pearson Education Inc. 336 p.

Angoff WH. 1971. Scales, norms and equivalent scores. In: Thorndike RL (Editor). Educational Measurement. 2nd Ed. Washington, DC: American Council on Education. 508 p.

Bergman EM, Prince KJ, Drukker J, van der Vleuten CP, Scherpbier AJ. 2008. How much anatomy is enough? Anat Sci Educ 1:184–188.

Biggs JB, Tang C. 2011. Teaching for Quality Learning at University. 4th Ed. Maidenhead, UK: Open University Press. 480 p.

Böckers A, Mayer C, Böckers TB. 2014. Does learning in clinical context in anatomical sciences improve examination results, learning motivation, or learning orientation? Anat Sci Educ 7:3–11.

Bruner J. 1960. The Process of Education. 1st Ed. Cambridge, MA: Harvard University Press. 97 p.

Cecot T, Smith CF, Border S. 2013. Blending the virtual and physical teaching environment: The potential role of the anatomy e-booklet. *J Anat* 224:248.

Chirculescu ARM, Chirculescu M, Moris JF. 2007. Anatomical teaching for medical students from the perspective of European Union enlargement. *Eur J Anat* 11:63–65.

Clough RW, Lehr RP. 1996. Testing knowledge of human gross anatomy in medical school: an applied contextual-learning theory method. *Clin Anat* 9:263–268.

Cohen-Schotanus J, van der Vleuten CP. 2010. A standard setting method with best performing students as point of reference: Practical and affordable. *Med Teach* 32:154–160.

Crossley J, Humphris G, Jolly B. 2002. Assessing health professionals. *Med Educ* 36:800–804.

De Gruijter DM. 1985. Compromise models for establishing examination standards. *J Educ Meas* 22:263–269.

Drake RL. 1999. Anatomy in a changing medical curriculum. *Kaibogaku Zasshi* 74:487–490.

Duffield KE, Spencer JA. 2002. A survey of medical students' views about the purposes and fairness of assessment. *Med Educ* 36:879–886.

Ebel RL. 1979. Essentials of Educational Measurement. 3rd Ed. Upper Saddle River, NJ: Prentice Hall Inc. 388 p.

Fabrizio PA. 2013. Oral anatomy laboratory examinations in a physical therapy program. *Anat Sci Educ* 6:271–276.

Hafferty FW. 1998. Beyond curriculum reform: Confronting medicine's hidden curriculum. *Acad Med* 73:403–407.

Heylings DJ. 2002. Anatomy 1999-2000: The curriculum, who teaches it and how? *Med Educ* 36:702–710.

Jalili M, Hejri SM, Norcini JJ. 2011. Comparison of two methods of standard setting: The performance of the three-level Angoff method. *Med Educ* 45:1199–1208.

McHanwell S, Atkinson M, Davies DC, Dyball R, Morris J, Ockleford C, Parkin I, Standring S, Whiten S, Wilton J. 2007. A core syllabus in anatomy—adding common sense to need to know. *Eur J Anat* 11:S3–S18.

Meyer JHF, Land R. 2003. Threshold concepts and troublesome knowledge: Linkages to ways of thinking and practicing within the disciplines. In: Rust C (Editor). *Improving Student Learning: Theory and Practice Ten Years On*. 1st Ed. Oxford, UK: Oxford Centre for Staff and Learning Development (OCSLD) Inc. p 1–16.

Norcini JJ. 2003. Setting standards on educational tests. *Med Educ* 37:464–469.

Rowland S, Ahmed K, Davies DC, Ashrafian H, Patel V, Darzi A, Paraskeva PA, Athanasiou T. 2011. Assessment of anatomical knowledge for clinical practice: Perceptions of clinicians and students. *Surg Radiol Anat* 33:263–269.

Shaibah HS, van der Vleuten CP. 2013. The validity of multiple choice practical examinations as an alternative to traditional free response examination formats in gross anatomy. *Anat Sci Educ* 6:149–156.

Smith CF, Martinez-Álvarez C, McHanwell S. 2014. The context of learning anatomy: Does it make a difference? *J Anat* 224:270–278.

Smith CF, Mathias HS. 2010. Medical students' approaches to learning anatomy: Students' experiences and relations to the learning environment. *Clin Anat* 23:106–114.

Smith CF, Mathias HS. 2011. What impact does anatomy education have on clinical practice? *Clin Anat* 24:113–119.

Yaqinuddin A, Zafar M, Ikram MF, Ganguly P. 2013. What is an objective structured practical examination in anatomy? *Anat Sci Educ* 6:125–133.

Zhang G, Fenderson BA, Schmidt RR, Veloski JJ. 2013. Equivalence of student's scores on timed and untimed anatomy practical examinations. *Anat Sci Educ* 6:281–285.

FIGURE LEGENDS

Figure 1

University of Southampton Faculty of Medicine curriculum map showing the first two years of the program.

Figure 2

Photograph of Dr Smith helping students during a practical session in the Centre for Learning Anatomical Sciences Laboratory, note the students own additions to her workbook.

Figure 3

Diagrammatic representation of the laboratory room layout of the Integrated Anatomy Practical Paper (IAPP) examination.

Figure 4

Box and Whisker Plot graph of Semester 4 Integrated Anatomy Practical Paper (IAPP) results.

Figure 5

Graph representing the Integrated Anatomy Practical Paper (IAPP) fail rates between 2009-2013.

It is notable that the rise in fail rates 2009 – 2011 and then the improvement in 2012 occur in both years therefore this is not a cohort effect:

Semester 1& 2 2009 and 3 & 4 2010 are the same cohort but there is a consistent rise between 2009 and 2010 the same is true following the same cohort 2010 – 11.

However the cohort starting year 1 in 2011 show improvement by their year 2 in 2012 particularly in the summer IAPP (Semester 4).

Figure 6

Graph illustrating the mean IAPP results between 2009-2013.

TABLES

Table 1. Details of Bachelor of Medicine cohort sample sizes

Cohort	Year 1		Year 2	
	Semester 1	Semester 2	Semester 3	Semester 4
2011-2012	224	222	201	201
2012-2013	199	197	219	220

Table 2. Sample questions used in the 2012 Semester 4 Integrated Anatomy

Practical Paper (IAPP)

Discipline	Resources	Question	Answer and Marks (m)	Taxonomy rating	Ebel's Standard Setting
Anatomy	Cadaveric specimen showing the posterior abdominal wall	Pain from a renal stone in B is referred to which regions?	Lumbar and inguinal region (2m) T10-L1 (2m) Rule: 1m if incomplete 0m marks if over	3	2m Easy 1m Easy
Anatomy	Sagittal section of a cadaveric specimen	If this region was punctured which part of the peritoneal cavity would be entered?	Rectouterine Pouch/Pouch of Douglas (2m)	3	2m Moderate 1m Moderate
Physiology	Model of the male genital system	Give two functions of the secretion from this (A) structure.	Alkaline to counteract acid vagina; help maintain sperm; energy for sperm; nutrition for sperm; prevents clotting; transport sperm; aids motility; antibacterial: citric acid; phosphatases, zinc. Any 2, 1 m each 2m maximum	2	2m Easy 1m Easy
Pharmacology	Sagittal section of a cadaveric specimen	Name two drugs (or drug classes) used to inhibit contractility at point A in premature labor (24-33 weeks gestation).	Beta-2 adrenoceptor agonist (e.g. salbutamol, terbutaline) (1m); calcium channel antagonist (e.g. nifedipine) (1m); MgSO ₄ (1m); NSAID (e.g. indometacin) or prostaglandin inhibitors (1m); oxytocin receptor antagonist (e.g. atosiban) (1m).	4	2m Easy 1m Easy
Pathology	Photomicrograph	From the photomicrograph state the most likely diagnosis.	Diabetes (mellitus) (2m); glomerulonephritis (1m); glomerulosclerosis (0m).	4	2m Difficult 1m Difficult
Applied Clinical	Fetal skull	State one clinical sign that may be seen at A in a living baby.	Raised intracranial pressure (2m); dehydration (2m); hydrocephalus (2m); meningitis (2m).	3	2m Moderate 1m Moderate

Table 3. The Integrated Anatomy Practical Paper, Cronbach Alpha Coefficients.

Cohort	Year 1		Year 2	
	Semester 1	Semester 2	Semester 3	Semester 4
2011-2012	0.868	0.873	0.87	0.80
2012-2013	0.923	0.871	0.88	0.84

Note: Coefficient Alpha ≥ 0.8 represents a strong reliability

Table 4. Blooms Taxonomy rating for Semester 4 Spotter Examination and Integrated Anatomy Practical Paper (IAPP) Examination questions (2009-2013).

Blooms Taxonomy Ratings	Old Spotter Main 2009 N (%)	Old Spotter Supp 2009 N (%)	Transition between Spotter and IAPP Main 2010 N (%)	Transition between Spotter and IAPP Supp 2010 N (%)	IAPP Exam Main 2011 N (%)	IAPP Exam Supp 2011 N (%)	IAPP Exam Main 2012 N (%)	IAPP Exam Supp 2012 N (%)	IAPP Exam Main 2013 N (%)
Level 1 Remembering	43 (72)	46 (77)	30 (50)	27 (46)	24 (40)	27 (46)	27 (46)	22 (37)	27 (46)
Level 2 Understanding	14 (23)	13 (22)	25 (42)	26 (43)	21 (36)	23 (38)	17 (28)	23 (38)	19 (30)
Level 3 Applying	3 (5)	1 (1)	3 (5)	4 (7)	12 (20)	6 (10)	11 (18)	10 (17)	10 (17)
Level 4 Analyzing	0	0	2 (3)	2 (3)	2 (3)	3 (5)	5 (8)	4 (7)	4 (7)
Level 5 Evaluating	0	0	0	1 (1)	1 (1)	1 (1)	0	1 (1)	0
Level 6 Creating	0	0	0	0	0	0	0	0	0
Total	60 (100)	60 (100)	60 (100)	60 (100)	60 (100)	60 (100)	60 (100)	60 (100)	60 (100)

Supp. Supplementary/Repeated Examinations. The number of questions (N) within each Blooms taxonomy level is also expressed as a percent of the total number of questions within the spotter or IAPP examination.